文章编号 1004-924X(2017)04-1060-10

结合位姿约束与轨迹寻优的人体姿态估计

李庆武[1,2*],席淑雅[1],王 恬[1],马云鹏[1],周亮基[1]
(1.河海大学 物联网工程学院,江苏 常州 213022;
2.常州市传感网与环境感知重点实验室,江苏 常州 213022)

摘要:基于混合部件模型的人体姿态估计方法忽视了人体结构的对称位姿约束关系,从而导致对称部件容易被重复检测、人体姿态估计准确率较低,为此,提出一种基于位姿约束与轨迹寻优的姿态估计新方法。首先估计人体单部件和对称部件在单帧图像中的多个合理位置,利用对称部件之间的位姿约束关系构建标识部件。然后根据单部件和标识部件各自的目标优化函数,通过动态规划算法反复迭代获得初始轨迹候选集,再结合轨迹的全局特征剔除检测得分较低的运动轨迹。最后引入树形合约模型,联系时空上下文信息,准确求解出视频序列光滑且兼容的最优轨迹。在 N-best、Outdoor Pose 和 Scene 数据集中的实验结果表明,对于存在背景复杂、运动模糊、部件遮挡等问题的视频序列中,该方法平均姿态估计准确率达 87%以上,有效减少了对称部件的误判,提高了视频中人体姿态估计的准确率。

关 键 词:人体姿态估计;混合部件模型;位姿约束;最优轨迹
中图分类号:TP391 **文献标识码**:A **doi**:10.3788/OPE.20172504.1060

Human pose estimation based on configuration constraints and trajectory optimization

LI Qing-wu[1,2*], XI Shu-ya[1], WANG Tian[1], MA Yun-peng[1], ZHOU Liang-ji[1]

(1. *College of Internet of Things Engineering, Hohai University, Changzhou* 213022, *China*;
2. *Changzhou Key Laboratory of Sensor Networks and Environmental Sensing, Changzhou* 213022, *China*)
* *Corresponding author, E-mail: hiqw@hhuc.edu.cn*

Abstract: Because of ignoring the configuration constraints between symmetric body parts, the human pose estimation methods based on mixtures of parts may lead to a repetitive detection of symmetrical body parts and a low pose estimation accuracy. Therefore, a kind of new pose estimation method on the basis of pose constraint and trajectory optimization was put forward. Firstly, numerous reasonable locations of single part and symmetric parts of human in single-frame image should be estimated, and identification part should be constructed by utilizing pose constraint relationship among symmetric parts. Then initial trajectory candidates set shall be gained through repeated iteration of dynamic programming algorithm according to respective target optimization function of single part and identification part. Movement trajectory with relatively low detection score was removed by combining with global feature of trajectory. Finally, smooth and compatible optimal trajectory of video sequence was

收稿日期:2016-11-24;**修订日期**:2017-01-16.
基金项目:国家自然科学基金(41301448);江苏省重点研发计划(BE2016071)

correctly solved by introducing tree-based contract model and combining with contextual spatio-temporal information. Experimental result in N-best, Outdoor Pose and Scene dataset shows that in video sequence with complex background, blur movement and part blocking problems, average pose estimation accuracy of proposed method is greater than 87%, which reduces erroneous judgment of symmetric parts effectively and improves human pose estimation accuracy in video.

Key words: Human pose estimation; mixed parts model; pose constraint; optimal trajectory

1　引　言

随着视频图像采集设备的普及，计算机视觉技术已越来越多地用于对现实生活中获取的图像和视频进行高效、准确、及时的处理[1-4]。作为计算机视觉领域的研究热点，人体姿态估计在视频图像检索、人机交互、智能视频监控等领域具有重要的研究价值[5-6]。近年来，有关人体姿态估计的研究工作大都集中在单帧图像领域，且已经取得了一定的进展[7-11]。而对于视频中的人体姿态估计的研究工作却进展缓慢。

人体姿态估计方法有基于模型的估计和无模型估计两种，其中基于部件模型的方法是当前姿态估计研究的热点[12-15]。但是该模型忽视了人体对称部件之间的位姿约束关系[16-17]，因而容易在人体对称部件上产生重复检测问题。此外，视频中的人体姿态估计通常采用图优化方法，但是由于序列图像中的人体部件在时间和空间上存在逻辑上的多圈图结构[18]，传统图优化方法在优化姿态候选集时存在 NP-hard(Non-deterministic polynomial hard)问题[19]，且通常只能采取近似求解的策略获得接近最大后验概率(MAP)解的次优解[18]。

为了提高视频中人体姿态估计的准确率，文献[20]把原始部件模型分解为众多可进行高效准确求解的树形子模型；文献[16]通过引入人体对称部件之间存在的位姿约束关系，完善了原始部件模型；文献[21]提出一种在肘腕部引入光流约束的上半身人体姿态估计算法，对肢体采用“先拆分，后重组”方法近似求解；文献[18]则提出跟踪基于选择的策略，并利用连接树算法准确求解，但该算法难以扩展到视频中的全身人体姿态估计。上述方法均具有较强的前瞻性，但都不能在高效准确求解和充分利用人体对称部件之间的位姿约束关系方面取得平衡。

本文基于混合部件模型[8]提出了一种基于位姿约束与轨迹寻优的姿态估计方法，并将其用于视频中的全身人体姿态估计。该方法能够对人体对称部件之间的位姿约束关系进行合理建模，利用树形图结构的高效推理算法求解出光滑且兼容的最优轨迹。实验部分，在几组典型测试序列上测试了本算法与目前主流的视频姿态估计算法的性能。

2　混合部件模型

混合部件模型(如图 1 所示)使用多个部件类型模板表示人体的每一部分，从而提升了单帧图像上人体姿态估计的精度[8]。但该模型忽视了人体结构中存在的对称位姿约束关系，容易产生对称部位被重复检测的问题。

本文以文献[8]提出的混合部件模型为基础模型，其中模型组件个数 $K=26$，局部混合类型 $T=4$。该模型把人体定义为一个无向图 $G=(V,E)$，其中 $V=\{v_i|_{i=1}^{K}\}$ 表示无向图顶点，E 表示无向图中所有相连部位构成的边集。假定所有人体部件在图像 I 中的位置集合为 $Z=\{z_i|_{i=1}^{K}\}$，$\theta_i=\{v_i,z_i|v_i\in V,z_i\in Z\}$ 表示各部件在图像 I 中的一种状态配置，则人体姿态估计即是一个确定 $\theta=\{\theta_i|_{i=1}^{V}\}$ 合理取值的过程。

若用外观特征项 $\Phi(\theta_i)$ 表示顶点 v_i 在位置 z_i 处的概率，变形评分项 $\psi(\theta_i,\theta_j)$ 表示每一对顶点 v_i 和 v_j 分别在位置 z_i 和 z_j 处相互间的兼容性，则姿态估计问题可通过求解公式(1)中评分函数的最大值 $S^*=\mathrm{argmax}_\theta(S(\theta))$ 来解决，即确定每个顶点的最佳位置，同时保证顶点之间具有良好的兼容性。

$$S(\theta)=\sum_{\theta_i\in\theta}\Phi(\theta_i)+\sum_{(\theta_i,\theta_j)\in E}\psi(\theta_i,\theta_j),\quad(1)$$

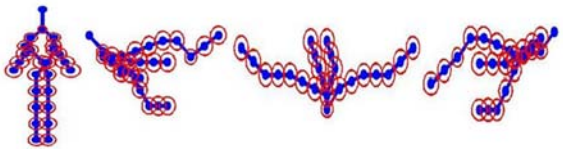


图 1 混合部件模型

Fig. 1 Mixture parts model

3 人体姿态估计

本文在混合部件模型的基础上引入了人体对称部件之间的位姿约束关系，并在姿态推理阶段引入轨迹寻优策略，从而使得视频中的人体姿态估计问题转变成基于树形图进行多阶段估计优化的问题。本文人体姿态估计的步骤如图 2(彩图见期刊电子版)所示。

(1) 首先利用 DivMBest 算法[19]，在视频序列的每一帧中产生多个不完全重叠且差异明显的姿态估计，如图 2(a)所示。

(2) 然后基于每个姿态的检测得分图，通过非极大值抑制(NMS)算法，获取每帧中各个部件初始的位置估计，再利用对称部件之间的位姿约束关系，合并脚踝等部件，构建概念上的“标识部件”，如图 2(b)所示；

(3) 利用(2)中得到的单部件和标识部件，构建视频中简化的人体马尔科夫(MRF)网络模型，消除人体各部件在帧内的空间中存在的多圈图结构，如图 2(c)所示，图中的每个顶点代表对应的人体部件；

(4) 接着根据单部件和标识部件各自的特点[16]，利用光流、区域 HOG 特征等设计单部件的目标优化函数，利用区域归一化的颜色直方图等设计标识部件的目标优化函数，通过动态规划算法获得各自的初始轨迹候选集。再结合轨迹的长度、加速度特征，在已获得的初始轨迹候选集中剔除综合检测得分较低的运动轨迹，得到进一步优化的轨迹候选集。其中，轨迹与轨迹间的空间关系如图 2(d)所示，图中的每个顶点代表对应人体部件的一条运动轨迹；

(5) 最后结合轨迹之间的兼容性特征，利用树形的合约模型求解出每个部件对应的最优轨迹，得到最终的人体姿态估计，如图 2(e)所示。

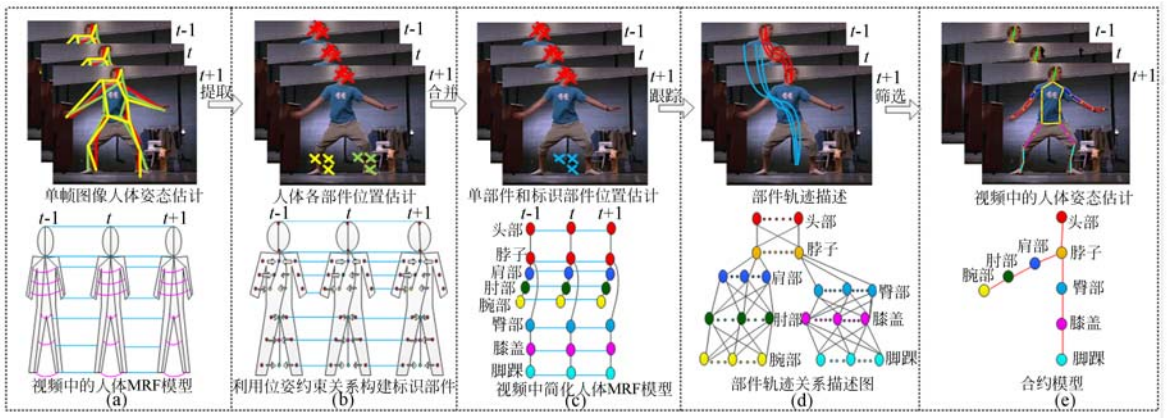


图 2 视频中的人体姿态估计

Fig. 2 Human pose estimation in video sequences

3.1 单帧图像的人体部件位置估计

本文首先采用 DivMBest 算法实现视频序列中单帧图像人体姿态的 M 个估计。然后从得到的 M 个最佳姿态估计当中，提取每个姿态各部件的位置坐标，这一过程基于检测得分图通过 NMS 算法采样获得。然后再用得到的左、右对称部件的位置组合表示相应标识部件的位置估计，而单部件的位置估计保持不变。根据文献[19]中的相关定义，检测得分函数 $\Phi_d(z)$ 的计算方法如公式(2)所示：

$$\Phi_d(z)=\sum_{i=1}^{V}\left(\Phi(z_i)+\sum_{m=1}^{M-1}\zeta_m[[z_i\neq z_i^m]]\right),\quad(2)$$

其中，z 表示部件在图像中的位置，$\Phi(z_i)$ 表示部件的外观特征得分，$\zeta=\{\zeta_m\mid m\in[M-1]\}$ 是一组拉格朗日数乘因子，其表示违反汉明约束的惩罚

权重，z_i 表示当前第 i 个部件位置的 MAP 解，z_i^m 表示当前第 i 个部件位置的第 m 个最优解，$[[z_i \neq z_i^m]]$ 表示当输入为真时，$[[\cdot]]$ 输出结果为 1，否则为 0。

3.2　部件运动轨迹描述

人体部件由两类组件构成，一类是头部、脖子等单部件，另一类是肘部、脚踝等的对称部件[16]，下面分别针对这两种组件采用不同的目标优化函数获得各自的初始轨迹候选集。

3.2.1　单部件运动轨迹描述

为了保证每条轨迹上的连接部件都具有较高的检测得分，本文从外观特征和相似性度量两个方面评析每个连接的单部件，采用 HOG 特征描述部件的外观特征，卡方距离进行相似性度量，并利用光流获取序列图像中运动显著的区域。其中，单部件的外观特征检测项用 $\Phi_d(z)$ 表示，它的计算方法见公式(2)，相似性度量项用 $\Psi_d(z^t, z^{t+1})$ 表示，计算方法如公式(3)所示：

$$\Psi_d(z^t, z^{t+1}) = \exp\left\{-\frac{\chi^2(T(\hat{z}^t), T(z^{t+1}))\cdot \| \hat{z}^t - z^{t+1} \|_2^2}{\sigma^2}\right\}, \tag{3}$$

其中，z^t 和 z^{t+1} 分别表示 t、$t+1$ 相邻两帧中部件所在的位置，$\chi^2(\cdot)$ 表示卡方距离，$T(z)$ 表示以位置 z 为中心的局部区域的 HOG 特征向量，$\hat{z}^t$ 表示通过光流预测的 z^t 在 $t+1$ 帧中所在的位置，σ 是模型参数。

该方法能够较好地保持相邻帧之间的姿态连续性，从而使姿态估计结果更加有效。根据上文求得的 $\Phi_d(z)$ 和 $\Psi_d(z^t, z^{t+1})$，利用公式(4)，通过动态规划算法可获得单部件得分最高的运动轨迹，然后移除当前选中的各帧中单部件的位置坐标，再次通过公式(4)获得下一条最优轨迹。重复这个过程，即可获得单部件得分较高的多条运动轨迹。

$$S_s(\theta^s) = \sum_{i=1}^{F} \Phi_d(\theta_i^s) + \sum_{i=1}^{F-1} \Psi_d(\theta_i^s, \theta_{i+1}^s), \tag{4}$$

其中，$\theta^s = \{\theta_i^s \mid_{i=1}^{F}\}$ 表示在帧数为 F 的视频序列中，单部件状态配置 θ_i^s 的集合，$S_s(\cdot)$ 为单部件运动轨迹的目标优化函数。

3.2.2　标识部件运动轨迹描述

每个标识部件的位置由对应的左、右对称部件 m 和 n 的位置坐标复合得到，本文用 $l=(m, n)$ 表示。标识部件能够对人体对称部件之间的位置互斥关系，以及外观的相似性进行建模，从而降低人体对称部件被重复检测的概率。为了更好地表示标识部件的外观特征，本文综合考虑检测得分置信度与两对称部件之间的兼容性，计算方法如下所示：

$$\Phi_c(l) = \frac{(\Phi_d(l.m) + \Phi_d(l.n))\cdot(\Lambda(l.m)^T \cdot \Lambda(l.n))}{1+\mathrm{e}^{-|l.m-l.n|/K}}, \tag{5}$$

其中，$\Phi_d(\cdot)$ 的计算方法同公式(2)，$l.m$ 和 $l.n$ 分别表示人体左、右对称部件，$\Lambda(z)$ 表示以 z 为中心的局部区域归一化后的颜色直方图，分母表示 S 形函数的倒数，用来惩罚重合的对称部件，K 用于控制惩罚程度。标识部件的变形评分项的计算方法如公式(6)所示：

$$\Psi_c(l^t, l^{t+1}) = \Psi_d(l.m^t, l.m^{t+1}) + \Psi_d(l.n^t, l.n^{t+1}), \tag{6}$$

其中，$\Psi_d(\cdot)$ 的计算方法同公式(3)。再利用公式(7)，通过 3.2.1 中介绍的方法即可获得标识部件得分较高的多条运动轨迹。

$$S_c(\theta^c) = \sum_{i=1}^{F} \Phi_c(\theta_i^c) + \sum_{i=1}^{F-1} \Psi_c(\theta_i^c, \theta_{i+1}^c), \tag{7}$$

其中，$\theta^c = \{\theta_i^c \mid_{i=1}^{F}\}$ 表示在帧数为 F 的视频序列中，标识部件状态配置为 θ_i^c 的集合，$S_c(\cdot)$ 为标识部件运动轨迹的目标优化函数。

3.2.3　基于全局特征的轨迹选择

全局特征能够很好地解决实际生活中由于引入相机噪声等因素造成的局部信息模糊问题。为了不增加姿态估计推理过程的复杂度，同时很好地利用轨迹的多种全局特征，本文仅在 3.2.1 和 3.2.2 中产生的多条单部件和标识部件运动轨迹的基础上使用全局特征进行轨迹筛选。

3.2.3.1　基于轨迹长度的全局特征

部件的运动轨迹如图 3 所示，每个包围盒 u 代表检测到的部件的一个位置，对所在帧中的位置按照时间顺序排列，数字为顺序号。本文用 λ 表示部件的一条运动轨迹，$TL(\lambda)$ 表示轨迹的总长度，如公式(8)所示：

$$TL(\lambda) = \sum_{u\in(i,j)} l(u), \tag{8}$$

其中，i、j 表示属于同一条运动轨迹的两个包围盒，$l(u)$ 表示两包围盒间的轨迹长度。部件的运动轨迹如图 3(彩图见期刊电子版)所示，红色轨迹和包围盒表示应舍弃的检测结果，蓝色轨迹表示应保留的检测结果。即两包围盒间的轨迹长度

越长，$TL(\lambda)$的值越大，该轨迹的置信度也就越高。

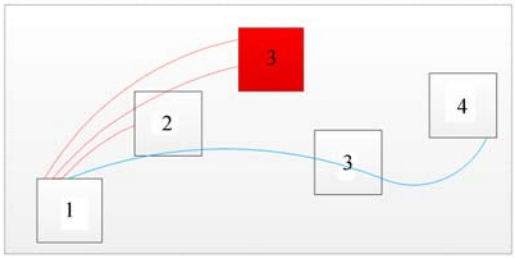


图 3　部件的运动轨迹

Fig. 3　Motion trajectories of body parts

3.2.3.2　基于轨迹加速度的全局特征

轨迹的光滑程度可以用加速度特征来衡量，加速度越小，轨迹越趋于光滑。本文用 A 来表示视频序列第 t 帧中的加速度特征，它的计算方法如公式(9)所示：

$$A(i,j,k;t)=\| z_i+z_k-2z_j \|, \tag{9}$$

其中，i,j,k 表示在 $t-1$、t 及 $t+1$ 相邻 3 帧中检测到的同一部件的包围盒，z_i，z_k，z_j 表示部件包围盒所在位置。则对于整条运动轨迹 λ，若模型参数用 ω 表示，则它的加速度特征 $TA(\cdot)$的计算方法如公式(10)所示：

$$TA(\lambda)=\exp\left(-\frac{\sum_t A(i,j,k;t)}{\omega}\right), \tag{10}$$

$TA(\cdot)$的值越小，则轨迹的光滑程度和置信度也越高。综合公式(8)和公式(10)可知，基于全局特征的轨迹优化目标函数如下所示：

$$S(\hat{\lambda})=S(\lambda)+TL(\lambda)+TA(\lambda), \tag{11}$$

其中，$S(\lambda)$为利用公式(4)或者公式(7)求得的多条运动轨迹的检测得分。$S(\hat{\lambda})$的值越大，则轨迹的置信度和光滑性越好。利用这个特征，依次剔除单部件和标识部件得分较低的运动轨迹。

3.3　基于轨迹寻优的人体姿态估计

为了确保人体各部件对应的轨迹之间具有良好的兼容特性，本文提出通过构建合约模型来选择最优轨迹的求解方式，如图 2(e)所示。该模型能够有效地利用树形图的高效求解策略，同时消除帧间存在的多圈图结构。其中，每个轨迹顶点的外观特征项得分由公式(11)确定，这里用 Φ_X 表示。根据文献[8]中对 $\omega_{i,j}$和 $\psi(\cdot)$的相关定义知，顶点与顶点之间的相对位置得分，也即两顶点之间的变形花费评分项 $\psi_d(p_i,q_j)=\omega_{i,j}\cdot\psi(p_i-q_j)$。则根据图 2(e)中顶点之间的边集类型，结合时空上下文，可将顶点与顶点之间的变形花费分为以下 3 种情况。

(1) 任意两单部件运动轨迹之间的连接：

$$\Psi_X(\lambda^s,\tau^s)=\sum_{i=1}^{F}\psi_d(\lambda_i^s,\tau_i^s), \tag{12}$$

其中，$\lambda^s=\{\lambda_i^s|_{i=1}^F\}$，$\tau^s=\{\tau_i^s|_{i=1}^F\}$表示任意相邻的两组单部件的运动轨迹。

(2) 任意单部件运动轨迹与标识部件运动轨迹之间的连接：

$$\Psi_X(\lambda^s,\tau^c)=\sum_{i=1}^{F}(\Psi_d(\lambda_i^s,\tau_i^c.m)+\Psi_d(\lambda_i^s,\tau_i^c.n)), \tag{13}$$

其中，$\lambda^s=\{\lambda_i^s|_{i=1}^F\}$，$\tau^c=\{\tau_i^c|_{i=1}^F\}$表示任意相邻的一组单部件和一组标识部件的运动轨迹。

(3) 任意两标识部件运动轨迹之间的连接：

$$\Psi_X(\lambda^c,\tau^c)=\sum_{i=1}^{F}(\Psi_d(\lambda_i^c.m,\tau_i^c.m)+\Psi_d(\lambda_i^c.n,\tau_i^c.n)), \tag{14}$$

其中，$\lambda^c=\{\lambda_i^c|_{i=1}^F\}$，$\tau^c=\{\tau_i^c|_{i=1}^F\}$表示任意两组相邻的标识部件的运动轨迹。则基于轨迹寻优策略的人体姿态估计的最终优化目标函数为：

$$S_X(\lambda^X)=\sum_{v_i^c\in V}\Phi_X(\lambda_i^X)+\sum_{(v_i^X,v_j^X)\in E}\Psi_X(\lambda_i^X,\lambda_j^X), \tag{15}$$

其中，$\lambda^X=\{\lambda_i^X|_{i=1}^v\}$表示任意一组满足图 2(e)中约束关系的轨迹顶点，接着通过动态规划算法高效、准确地求解公式(15)中目标函数的最大值 $\lambda^{X^*}=\mathrm{argmax}_{\lambda^X}(S(\lambda^X))$，由此即可获得视频序列中的最优轨迹，得到人体每个部件在视频序列每帧中的最终位置。

4　实验与结果分析

本文算法的程序代码使用 Matlab 与 C++混编，在 Matlab2012a 版本上编写；算法的运行环境为 CPU：Intel(R) Core(TM) i7-4719HQ CPU @2.5GHz；内存：4.00GB 的笔记本。

4.1　测试数据集及评价指标

本文采用 3 个数据集对算法进行评估，分别为：(1) N-best 数据集。该数据集由 Park 等人在文献[14]中提出，共包含 4 段视频序列，约 600 帧已标注人体各部件真实值的图像。该数据集的图

像中包含丰富的运动模糊和姿态变化。(2) Outdoor Pose 数据集。该数据集由 Ramakrishna 等人在文献[16]中提出，共包含 6 段视频序列，约 1 000帧已标注人体各部件真实值的图像。该数据集的图像中包含丰富的人体自遮挡。N-best 数据集和 Outdoor Pose 数据集在姿态估计领域均被普遍认可。(3) Scene 数据集。该数据集由本文提出，主要用来测试算法在具有较强背景噪声干扰情况下的姿态估计效果，共包含 20 段取自多部电影的视频序列，约 600 帧图像。其中，每帧图像的人体各部件的真实值都经过手工标定。

本文采用 PCP 评价标准来评估算法对于人体各部件的估计准确度[22]。该评价标准是人体姿态估计中公认的评价标准，它不设定明确的像素值作为阈值，从而使评价结果不受图像中人物尺寸的像素值或图片大小的影响，能够有效评价不同视频序列的姿态估计精度。该评价指标给出了模型正确定位的人体部件百分比，可以通过下式来计算：

$$PCP=\frac{n_{\text{Correct}}}{n_{\text{Limbs}}\times n_{\text{Images}}},\tag{16}$$

其中 n_{Correct} 表示所有图像中被正确定位的部件的总数，n_{Limbs} 表示每幅图像中所要定位的人体部件的数量，n_{Images} 表示测试集合的大小或正确定位人体位置的图像总数。规定当估计的所有部件端点到其对应真实值端点的距离小于部件长度的一半时，则认为该部件被正确定位。其中，该评价标准的最大值为 1，其值越大表示对人体各部件的估计准确度越高。

4.2　实验结果与分析

实验中每次处理相邻的 10 帧视频序列，3.1 节中的人体标识部件估计数目设置为 30，3.2.1 和 3.2.2 节中产生的部件轨迹数目设置为 20，3.2.3 节保留初始轨迹候选集中得分较高的前 10 条运动轨迹。公式(2)中的参数 ζ 及公式(10)中的参数 ω 均在 Leeds Sports Pose Dataset 训练数据集中利用网格搜索法获得，公式(3)中的参数 σ 和公式(5)中的参数 K 取 3.1 节中得到的 M-best 人体姿态估计高度中值的 10%。

表 1 给出了在 N-best、Outdoor Pose 和 Scene 3 个数据集上，本文方法与其他文献方法的对比实验结果。其中，文献[21]本身是一种用作上半身的人体姿态估计方法，本文在原文公开的代码基础上，把它扩展成了一个全身的人体姿态估计方法。文献[16]没有公开的源码，所以本文使用其原文中提供的实验数据用作对比实验。从表 1 中可以看出，与其他相关的前沿算法相比，本文方法对于人体各部件的估计准确率较高，并且对胳膊下端和腿下端等灵活部件的估计准确率有显著提高。由表 1 可知，在 N-best 数据集中，本文方法对于人体各部件的平均估计准确率较文献[14]、文献[16]和文献[21]分别提升了 7%、8%和 15%；在 Outdoor Pose 数据集中，本文方法对于人体各部件的平均估计准确率较文献[14]、文献[16]和文献[21]分别提升了 10%、3%和 7%；在 Scene 数据集中，本文方法对于人体各部件的平均估计准确率较文献[14]和文献[21]分别提升了 6%和 9%。

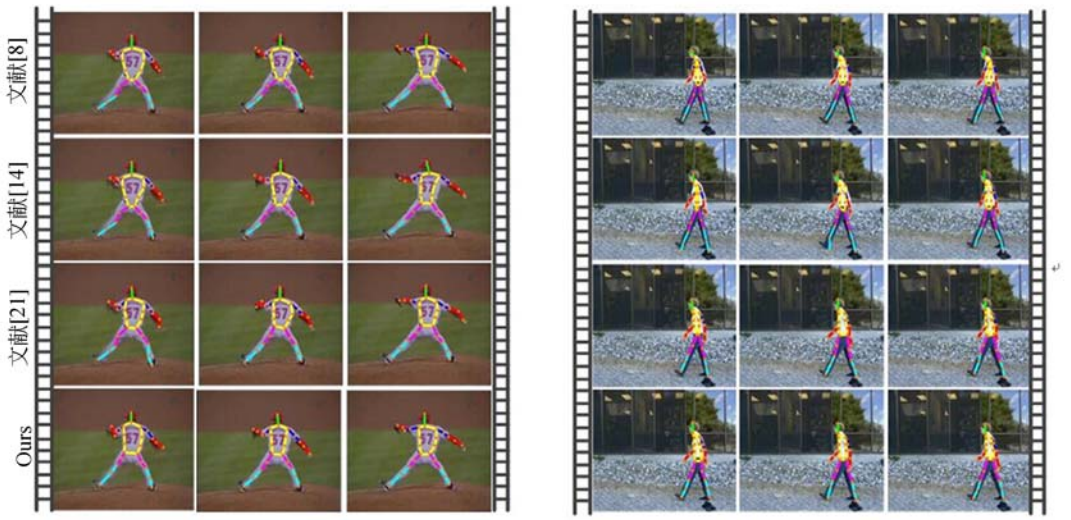
表 1　人体姿态估计准确度比较

Tab. 1　Accuracy comparison of human pose estimation algorithms

方 法	躯干	头 部	腿 上 端	腿 下 端	胳 膊 上 端	胳 膊 下 端	平 均
			N-best Dataset				
文 献[21]	1.0	1.0	0.91	0.90	0.67	0.40	0.81
文 献[16]	0.69	1.0	0.91	0.89	0.85	0.42	0.80
文 献[14]	0.61	1.0	0.86	0.84	0.66	0.41	0.73
本文方法	1.0	1.0	0.92	0.92	0.87	0.55	0.88
			Outdoor Pose Dataset				
文 献[21]	0.96	0.88	0.69	0.89	0.79	0.51	0.79
文 献[16]	0.86	0.99	0.95	0.96	0.86	0.52	0.86
文 献[14]	0.83	0.99	0.92	0.86	0.79	0.52	0.82
本文方法	0.97	0.99	0.97	0.96	0.87	0.61	0.89
			Scene Dataset				
文 献[14]	0.86	0.98	0.91	0.85	0.77	0.49	0.81
文 献[21]	0.79	0.96	0.87	0.82	0.75	0.49	0.78
本文方法	0.93	0.99	0.93	0.89	0.88	0.59	0.87

图 4,图 5 和图 6 分别为本文方法在 N-best、Outdoor Pose 和 Scene 数据集上的姿态估计结果图。在每个数据集中选取两组测试序列,每组测试序列的每一行为连续的 3 帧图像。同时,本文选取文献[8]的方法作为实验的参照标准。图 4(a)、图 4(b)分别为 baseball 和 walkstraight 测试序列中部分图片的实验结果,可见,图 4(a)中的人体四肢和躯干位置存在大量的运动模糊,图 4(b)中的人体右胳膊被躯干轻微遮挡,且在右脚的前端附近存在颜色相近的背景干扰物。实验结果表明,在图 4(a)中,文献[21]对于人体胳膊下端和躯干的估计准确率较本文方法明显偏低,而在 4(b)中,文献[14]和文献[21]还出现了将非肢体区域误检为肢体的情况。图 5(a)、图 5(b)分别为 HSP_CIC_bounce 和 VR_NSH_Kick 测试序列中部分图片的实验结果,前者图像中的人体左胳膊和躯干之间存在自遮挡问题,后者图像中的背景较为复杂。在图 5(a)中,与文献[8]、文献[14]和文献[21]相比,本文方法对于人体左前臂的估计准确率较高,而在图 5(b)中,4 种方法对于人体头部和躯干的估计准确率均较高,但文献[8]方法对于人体对称部件存在重复检测的问题。图 6(a)、图 6(b)分别为 Outdoor_winter 和 Outdoor_autumn 测试序列中部分图片的实验结果,图 6(a)中的人体衣物遮挡明显,图 6(b)中的背景区域极为复杂。在图 6(a)中,对于人体对称部件,文献[8]和文献[14]的方法均出现了不同程度的重复检测问题,而在图 6(b)中,文献[8]、文献[14]和文献[21]对于人体头部、躯干和腿下端的估计准确率明显比本文方法低。由此可得,本文方法有效地解决了基于混合部件模型的人体姿态估计方法中存在的对称部件被重复检测的问题,提高了现有方法的人体姿态估计准确度。

图 7 为本文方法在测试序列集中对一些姿态估计失败的样例。图 7(a)中的姿态均为稀有姿态,即该姿态在训练数据集中出现的次数较少或者不存在,导致基于特定数据集训练得到的方法对于这些稀有姿态的适应性较差,姿态估计准确率低。图 7(b)中存在严重的人体自遮挡及背景遮挡问题,导致被遮挡的人体部件无法被正确估计。

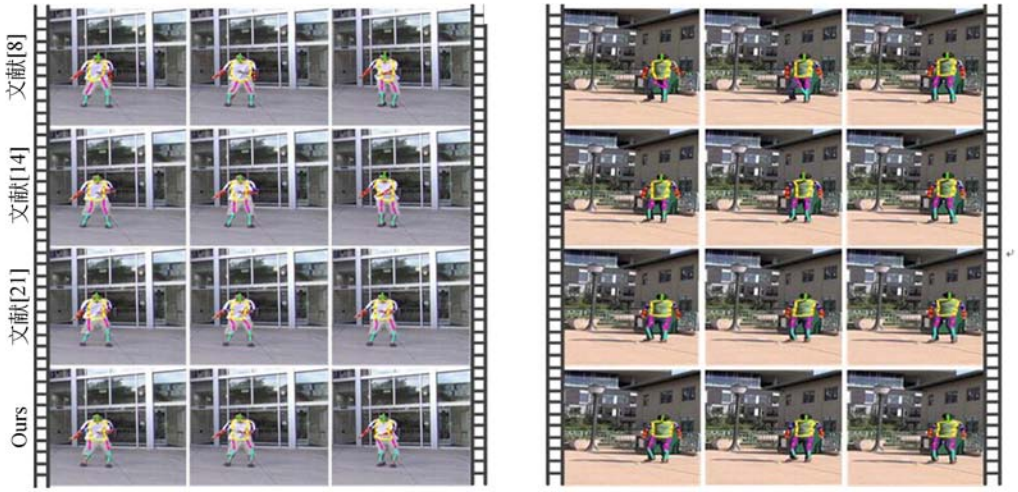


(a)Baseball 测试序列　　(b) Walkstraight 测试序列

(a) Baseball test sequences　　(b) Walkstraight test sequences

图 4　姿态估计结果图(N-best 数据集)

Fig. 4　Pose estimation results(N-best Datasets)



（a）HSP_CIC_bounce 测试序列　　（b）VR_NSH_Kick 测试序列

（a）HSP_CIC_bounce test sequences　　（b）VR_NSH_kick test sequences

图 5　姿态估计结果图（Outdoor Pose 数据集）

Fig. 5　Pose estimation results（Outdoor Pose Datasets）

（a）Outdoor_winter 测试序列　　（b）Outdoor_autumn 测试序列

（a）Outdoor_winter test sequences　　（b）Outdoor_autumn test sequences

图 6　姿态估计结果图（Scene 数据集）

Fig. 6　Pose estimation results（Scene Datasets）

（a）稀有姿态情况　　（b）遮挡情况

（a）Rare poses situation　　（b）Occlusion situation

图 7　姿态估计失败样例

Fig. 7　Failure cases of pose estimation

5 结　论

针对视频中人体姿态估计准确率低的问题，本文提出了一种基于位姿约束与轨迹寻优相结合的姿态估计方法。该方法利用构建的标识部件，对人体对称部件之间的位置互斥及外观相似性关系进行合理建模，从而较好地解决了对称部件被重复检测的问题。利用树形图结构的高效求解策略，分阶段对视频中的姿态估计问题进行优化，解决了NP-hard问题。实验结果表明，本文方法的平均姿态估计准确率达87%以上，相较于目前优秀的视频人体姿态估计算法，其能够更加准确地检测出人体的各个部件。

参考文献：

[1] 李静宇，刘艳滢，田睿，等. 视频监控系统中的概率模型单目标跟踪框架[J]. 光学 精密工程，2015，23(7)：2093-2099.
LI J Y，LIU Y Y，TIAN R，*et al*.. Probabilistic model single target tracking framework for video surveillance system[J]. *Opt. Precision Eng.*，2015，23(7)：2093-2099.(in Chinese)

[2] 李玉峰，李广泽，谷绍湖，等. 基于区域分块与尺度不变特征变换的图像拼接算法[J]. 光学 精密工程，2016，24(5)：1197-1205.
LI Y F，LI G Z，GU SH H，*et al*.. Image mosaic algorithm based on area blocking and SIFT[J]. *Opt. Precision Eng.*，2016，24(5)：1197-1205.(in Chinese)

[3] 胡梦婕，魏振忠，张广军. 基于对象性测度估计和霍夫森林的目标检测方法[J]. 红外与激光工程，2015，44(6)：1936-1941.
HU M J，WEI ZH ZH，ZHANG G J. Object detection method based on objectness estimation and Hough forest[J]. *Infrared and Laser Engineering*，2015，44(6)：1936-1941.(in Chinese)

[4] 杨磊，任龙，刘庆，等. 基于FPGA的大视场图像实时拼接技术的研究与实现[J]. 红外与激光工程，2015，44(6)：1929-1935.
YANG L，REN L，LIU Q，*et al*.. Research and implementation of large field image real-time mosaic technology based on FPGA[J]. *Infrared and Laser Engineering*，2015，44(6)：1929-1935.(in Chinese)

[5] 田国会，尹建芹，韩旭，等. 一种基于关节点信息的人体行为识别新方法[J]. 机器人，2014，36(3)：285-292.
TIAN G H，YIN J Q，HAN X，*et al*.. A novel human activity recognition method using joint points information[J]. *Robot*，2014，36(3)：285-292.(in Chinese)

[6] 韩贵金，朱虹. 基于HOG和颜色特征融合的人体姿态估计[J]. 模式识别与人工智能，2014，27(9)：769-777.
HAN G J，ZHU H. Human pose estimation based on fusion of HOG and color feature[J]. *PR&AI.*，2014，27(9)：769-777.(in Chinese)

[7] EICHNER M，MARIN-JIMENEZ M，ZISSERMAN A，*et al*.. 2D articulated human pose estimation and retrieval in (almost) unconstrained still images[J]. *International Journal of Computer Vision*，2012，99(2)：190-214.

[8] YANG Y，RAMANAN D. Articulated human detection with flexible mixtures of parts[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*，2013，35(12)：2878-2890.

[9] GKIOXARI G，HARIHRAN B，GIRSHICK R，*et al*.. Using k-poselets for detecting people and localizing their keypoints[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*，2014：3582-3589.

[10] LÓPEZ-QUINTERO M I，MARÍN-JIMÉNEZ M J，MUÑOZ-SALINAS R，*et al*.. Stereo pictorial structure for 2D articulated human pose estimation[J]. *Machine Vision and Applications*，2016，27(2)：157-174.

[11] SHOTTON J，SHARP T，KIPMAN A，*et al*.. Real-time human pose recognition in parts from single depth images[J]. *Communications of the ACM*，2013，56(1)：116-124.

[12] OUYANG W，CHU X，WANG X. Multi-source deep learning for human pose estimation[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*，2014：2337-2344.

[13] BUEHLER P，EVERINGHAM M，HUTTENLOCHER D P，*et al*.. Upper body detection and tracking in extended signing sequences[J]. *International Journal of Computer Vision*，2011，95

(2)：180-197.

[14] PARK D，RAMANAN D. N-best maximal decoders for part models[C]. *Proceedings of the IEEE International Conference on Computer Vision*，2011：2627-2634.

[15] 马淼，李贻斌. 基于多级动态模型的 2 维人体姿态估计[J]. 机器人，2016，38(5)：578-587.
MA M，LI Y B. 2D human pose estimation using multi-level dynamic model[J]. *Robot*，2016，38(5)：578-587.(in Chinese)

[16] RAMAKRISHNA V，KANADE T，SHEIKH Y. Tracking human pose by tracking symmetric parts[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*，2013：3728-3735.

[17] RAMAKRISHNA V，MUNOZ D，HEBERT M，*et al*.. Pose machines：articulated pose estimation via inference machines[C]. *European Conference on Computer Vision*，2014：33-47.

[18] TOKOLA R，CHOI W，SAVARESE S. Breaking the chain：liberation from the temporal Markov assumption for tracking human poses[C]. *Proceedings of the IEEE International Conference on Computer Vision*，2013：2424-2431.

[19] BATRA D，YADOLLAHPOUR P，GUZMAN-RIVERA A，*et al*.. Diverse M-best solutions in markov random fields[C]. *European Conference on Computer Vision*，*Springer Berlin Heidelberg*，2012：1-16.

[20] SAPP B，WEISS D，TASKAR B. Parsing human motion with stretchable models[C]. *Computer Vision and Pattern Recognition*，*IEEE*，2011：1281-1288.

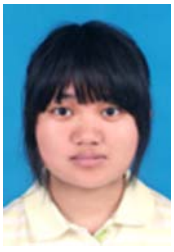
[21] CHERIAN A，MAIRAL J，ALAHARI K，*et al*.. Mixing body-part sequences for human pose estimation[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014：2361-2368.

[22] FERRARI V，MARIN-JIMENEZ M，ZISSERMAN A. Progressive search space reduction for human pose estimation[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*，2008：1-8.

通讯作者：

李庆武(1964－)，男，河南新乡人，博士、教授、博士生导师，1985 年于郑州大学获得学士学位，1990 年于西安电子科技大学获得硕士学位，2010 年于河海大学获得博士学位，主要研究方向为智能感知与图像处理。E-mail：liqw@hhuc.edu.cn

作者简介：



席淑雅(1993－)，女，河南商丘人，硕士研究生，2015 年于河海大学获得学士学位，主要研究方向为数字图像处理。E-mail：xishuya@hhu.edu.cn